

Safe AI Engineering Methodology

Dr. Elena Hoemann, Dr.-Ing. Sven Hallerbach (DLR)
Dr. Patrick Feifel (Opel)
Dr. Maximilian Grabowski, Fabian Szydelko (BAST)

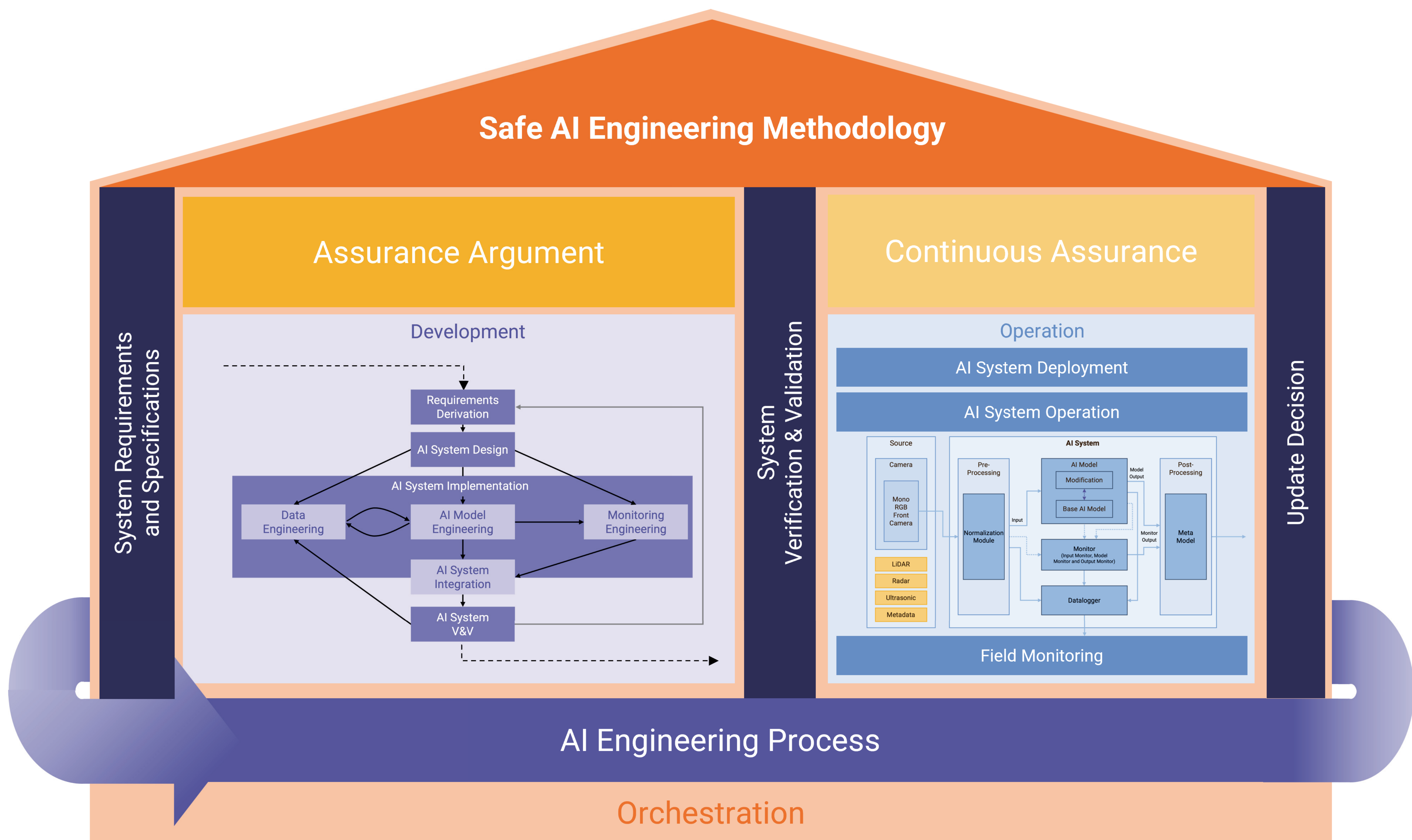


Figure 1: Safe AI engineering methodology (© DLR, Opel, BAST)

1 | New engineering concepts needed

The use of AI in automated driving provides new opportunities. However, it also carries a substantial risk: the so-called „black box“ nature of AI systems makes their decision-making process difficult to understand. This complicates the process of providing a traditional safety evaluation. Therefore, new engineering concepts are needed to use AI for safety-critical applications, such as automated driving.

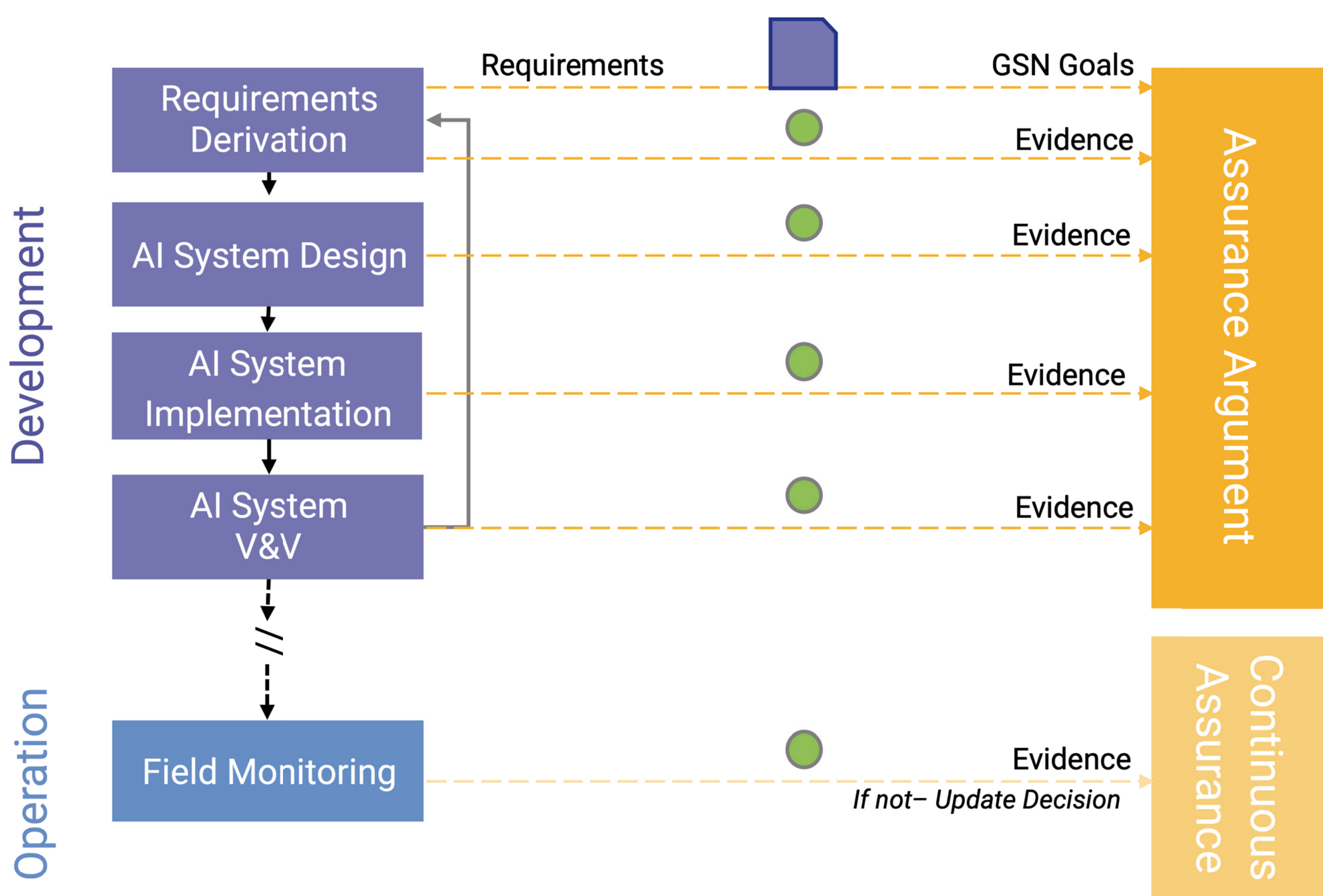


Figure 3: AI engineering process mapped to assurance argument (© DLR, Opel, BAST)

3 | Process Mapped to Assurance Argument

The interaction of process and assurance argument is depicted in Figure 3. The requirements are mapped to goals in the GSN structure and throughout the process evidence is collected to support that the requirements are fulfilled.

During operation, field monitoring continuously assesses whether the assurance argument is still valid. If not, an update decision can be triggered.

2 | The Methodology

The methodology is embedded into the context of the overarching system. System requirements and specifications are used as input to the AI engineering process. During the development of the AI system an assurance argument is built up in parallel. The process consists of four iterative steps: requirements derivation, design, implementation and V&V. The implementation is then split up into data, AI model and monitoring engineering, each triggering those process steps on a lower level. After AI system V&V, the V&V process at the system level continues.

After deployment, the operating AI system is depicted (Figure 1), including field monitoring to support update decisions.

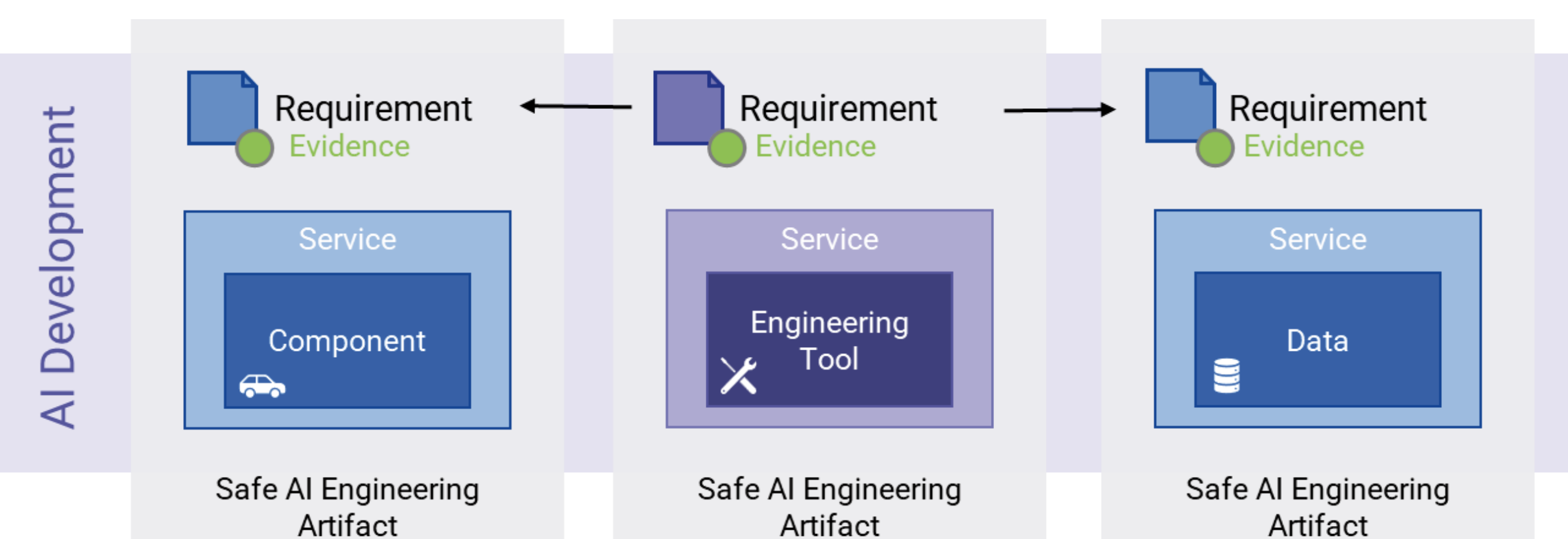


Figure 2: Safe AI engineering artifact (© DLR, Opel, BAST)

4 | The Safe AI Engineering Artifact

The AI engineering artifact always consists of a requirement, a service and evidence that the requirement is met. A service in this context can either be data or a component (blue, Figure 2), which will later be part of the operation phase, or an engineering tool helping to develop another component (purple, Figure 2).

5 | Conclusion

The methodology brings together the assurance argument with the engineering of data, model and monitors, as well as the integration in the embedded system.

For more information regarding the poster, contact:

elena.hoemann@dlr.de,
patrick.feifel@stellantis.com,
grabowski@bast.de

Safe AI Engineering is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.



[linkedin.com/company/ki-familie](https://www.linkedin.com/company/ki-familie)
www.safe-ai-engineering.de